Problem 1

Part A

Since the probability that $X_n = k$ for some $1 \le k \le n$ is uniform,

$$\mathbb{E}[X_n] = \sum_{k=1}^{n} \mathbb{P}[X_n = k] \cdot \omega(k) = \frac{1}{n}\sum_{k=1}^{n} \omega(k).$$

Rather than writing this sum in terms of the number of prime factors per given number, we can write this sum in terms of how many numbers a given prime p factors into between 1 and n. Thus

$$\sum_{k=1}^{n} \omega(k) = \sum_{\substack{p \le n \\ p \text{ prime}}} \sum_{\substack{1 \le k \le n \\ p|k}} 1.$$

The inner sum is simply how many numbers from 1 to n are divisible by some p, which is just $\lfloor \frac{n}{p} \rfloor$. Therefore

$$\mathbb{E}[X_n] = \frac{1}{n} \sum_{\substack{p \le n \\ p \text{ prime}}} \left\lfloor \frac{n}{p} \right\rfloor.$$

When $n \to \infty$, we have $\frac{1}{n}\lfloor \frac{n}{p} \rfloor = \frac{1}{p}$. Therefore for large n

$$\mathbb{E}[X_n] \approx \sum_{\substack{p \le n \\ p \text{ prime}}} \frac{1}{p}.$$

Thus by Merten's Theorem,

$$\mathbb{E}[X_n] = \log\log n + O(1).$$

Part B

Note that

$$\omega(X_n)^2 = \sum_{\substack{p \le n \\ p \text{ prime}}} \mathbb{1}_{p|X_n} + \sum_{\substack{p < q \le n \\ p \text{ prime}}} \mathbb{1}_{p|X_n} \mathbb{1}_{q|X_n}.$$

Taking the expectation of each term gives

$$\mathbb{E}\left[\sum_{\substack{p\le n\\ p\text{ prime}}} \mathbb{1}_{p|X_n}\right] = \sum_{\substack{p\le n\\ p\text{ prime}}} \frac{1}{p} = \log\log n + O(1)$$

and by independence of $\mathbb{1}_{p|X_n}$ and $\mathbb{1}_{q|X_n}$ since p and q are distinct,

$$\mathbb{E}\left[\sum_{\substack{p<q\le n\\ p\text{ prime}}} \mathbb{1}_{p|X_n}\mathbb{1}_{q|X_n}\right] = \sum_{\substack{p<q\le n\\ p\text{ prime}}} \frac{1}{pq}.$$

This sum across distinct primes can be rewritten as

$$\sum_{\substack{p<q\le n\\ p\text{ prime}}} \frac{1}{pq} = \left(\sum_{\substack{p\le n\\ p\text{ prime}}} \frac{1}{p}\right)^2 - \sum_{\substack{p\le n\\ p\text{ prime}}} \frac{1}{p^2}$$

which comes from the fact that the square of the sum of prime reciprocals will include all products $\frac{1}{pq}$ but will include all $\frac{1}{p^2}$ terms which must be removed. Note that

$$0 \le \sum_{\substack{p\le n\\ p\text{ prime}}} \frac{1}{p^2} \le \sum_{n=1}^{n} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

Thus the reciprocal prime square sum is $O(1)$. By Merten's theorem we have then

$$\sum_{\substack{p<q\le n\\ p\text{ prime}}} \frac{1}{pq} = (\log\log n)^2 + O(1).$$

By the definition of expectation we get in total

$$\begin{aligned}
\operatorname{Var}\omega(X_n) &= \mathbb{E}\left[\omega(X_n)^2\right] - \mathbb{E}[\omega(X_n)]^2 \\
&= (\log\log n)^2 + \log\log n - (\log\log n)^2 + O(1) \\
&= \log\log n + O(1)
\end{aligned}$$

Part C

Note that Chebyshev's inequality states

$$\mathbb{P}[|\omega(X_n) - \mathbb{E}[\omega(X_n)]| \geq a] \leq \frac{\operatorname{Var}\omega(X_n)}{a^2}.$$

Picking $a = t\sqrt{\operatorname{Var}\omega(X_n)}$ gives

$$\mathbb{P}\left[|\omega(X_n) - \mathbb{E}[\omega(X_n)]| \geq t\sqrt{\operatorname{Var}\omega(X_n)}\right] \leq \frac{1}{t_2}$$

which clearly then goes to 0 as $t \to \infty$.

Problem 2

Proof. Note that $\deg v \sim \operatorname{Bin}(n-1, p)$. Thus $\mathbb{E}[\deg v] = \frac{n-1}{2}$. For large n, $\frac{n-1}{2} \approx \frac{n}{2}$, thus $\mathbb{E}[\deg v] = \frac{n}{2}$. Using the Chernoff bound gives

$$\mathbb{P}[|\mathbb{E}[\deg v - \mathbb{E}[\deg v]]| \geq \delta\mathbb{E}[X]] \leq 2\exp\left(-\frac{\mathbb{E}[X]\delta^2}{3}\right).$$

Let $t = \sqrt{10n\log n}$ and choose $\delta = \frac{t}{\mathbb{E}[X]}$ which equals

$$\frac{2\sqrt{10n\log n}}{n} = 2\sqrt{\frac{10\log n}{n}}.$$

For large n, this is between 0 and 1 and hence is a valid choice of δ. Substituting into the right hand side of the bound gives

$$2\exp\left[-\left(\frac{n}{2}\right)\left(\frac{40\log n}{n}\right)\left(\frac{1}{3}\right)\right] = 2\exp\left[-\frac{20\log n}{3}\right] = 2n^{-\frac{20}{3}}.$$

Thus by applying the union bound,

$$\mathbb{P}\left[\exists v \text{ s.t. } |\deg v - \frac{n}{2}| \geq \sqrt{10n\log n}\right] = n\left(2n^{-\frac{20}{3}}\right) = 2n^{-\frac{17}{3}}$$

which goes to 0 as $n \to \infty$. ◇

Problem 3

Part A

Proof. Let $\mu = \mathbb{E}[X]$. Consider when it is the case that $|X - \mu| \geq |\mu|$. It is guaranteed to be true if $X = 0$. Therefore it is the case that

$$\mathbb{P}[X = 0] \leq \mathbb{P}[|X - \mu| \geq |\mu|].$$

Since $|\mu|$ is non-negative, by Chebyshev's inequality

$$\mathbb{P}[X = 0] \leq \mathbb{P}[|X - \mu| \geq |\mu|] \leq \frac{\operatorname{Var} X}{(|\mu|)^2} = \frac{\operatorname{Var} X}{\mu^2}.$$

Assume now that for a sequence of random variables X_n that $\lim_{n\to\infty} \frac{\operatorname{Var} X_n}{\mathbb{E}[X_n]^2} = 0$. Probabilities are non-negative meaning

$$0 \leq \mathbb{P}[X_n = 0] \leq \frac{\operatorname{Var} X_n}{\mathbb{E}[X_n]^2}.$$

Hence by the squeeze lemma it must be the case that as $n \longrightarrow \infty$, $\mathbb{P}[X_n = 0] = 0$. Thus we can conclude with high probability that $X_n \neq 0$. ◇

Part B

Proof. Let X denote the total number of 4-cliques in $G_{n,p}$. This can be expressed as a sum of indicator random variables $X_{i,j,k,l}$ which are 1 iff the vertices i, j, k, l are in a 4-clique together. Thus

$$X = \sum_{\substack{1 \leq i,j,k,l \leq n \\ i \neq j \neq k \neq l}} X_{i,j,k,l}.$$

By linearity of expectation,

$$\mathbb{E}[X] = \binom{n}{4} \cdot \mathbb{E}\left[X_{1,2,3,4}\right].$$

The probability that 4 vertices are in a clique is the product of the probabilities that each pair of vertices is connected, giving

$$\mathbb{P}\left[X_{1,2,3,4} = 1\right] = p^{\binom{4}{2}} = p^6.$$

Therefore

$$\mathbb{E}[X] = \binom{n}{4} \cdot p^6.$$

Taking $p \geq Cn^{-\frac{2}{3}}$ it follows for large C

$$\mathbb{E}[X] \geq \frac{n^4}{24}(C^6 n^4) = \frac{C^6}{24}.$$

This lower bound goes to infinity as both $n, C \to \infty$, thus $\mathbb{E}[X] \to \infty$.

$\diamond$

Part C

Using the expectation calculate from (B) and taking $p \leq \varepsilon n^{-\frac{2}{3}}$ it follows

$$\mathbb{E}[X] = \binom{n}{4} p^6 \leq \frac{n^4}{24}\left(\varepsilon n^{-\frac{2}{3}}\right)^6 = \frac{\varepsilon^6}{24}.$$

By Markov's inequality we then have as both $n \to \infty$ and $\varepsilon \to 0$ that

$$\mathbb{P}[X \geq 1] \leq \mathbb{E}[X] \leq \frac{\varepsilon^6}{24} \to 0.$$

Problem 4

Note that being uncorrelated implies independence, thus the random variables are pairwise independent.

Part A

By the definition of correlation,

$$\sigma_{X_1+X_2, X_3+X_4} = \frac{\operatorname{Cov}(X_1 + X_2, X_3 + X_4)}{\sqrt{\operatorname{Var}(X_1 + X_2)\operatorname{Var}(X_2 + X_3)}}.$$

First we find the covariance. Let μ_i be the expectation of the ith random variable. Then

$$\begin{aligned}\mathrm{Cov}(X_1+X_2, X_3+X_4) &= \mathbb{E}[(X_1+X_2)(X_3+X_4)] - \mathbb{E}[X_1+X_2]\mathbb{E}[X_3+X_4] \\ &= \mathbb{E}[X_1X_3+X_1X_4+X_2X_3+X_2X_4] - (\mu_1+\mu_2)(\mu_3+\mu_4) \\ &= \mu_1\mu_3+\mu_1\mu_4+\mu_2\mu_3+\mu_2\mu_4 \\ &= 0\end{aligned}$$

Since the covariance is 0, it follows that $\sigma_{X_1+X_2,X_3+X_4} = 0$.

Part B

We again find the covariance.

$$\begin{aligned}\mathrm{Cov}(X_1+X_2, X_2+X_3) &= \mathbb{E}[(X_1+X_2)(X_2+X_3)] - \mathbb{E}[X_1+X_2]\mathbb{E}[X_2+X_3] \\ &= \mathbb{E}\left[X_1X_2+X_1X_3+X_2^2+X_2X_3\right] - (\mu_1+\mu_2)(\mu_2+\mu_3) \\ &= \mathbb{E}[X_2]^2+\mu_1\mu_2+\mu_1\mu_3+\mu_2\mu_3 \\ &= \mathbb{E}[X_2]^2-\mu_2^2 \\ &= \mathrm{Var}\,X_2 = 1\end{aligned}$$

Since all variables are pairwise uncorrelated, $\mathrm{Var}(X_1+X_2) = \mathrm{Var}(X_2+X_3) = 2$. In total then

$$\sigma_{X_1+X_2,X_2+X_3} = \frac{1}{\sqrt{2}^2} = \frac{1}{2}.$$

Problem 5

Let N be the number of accidents in a week, and X_i the number of workers injured during the ith accident. The total number of worker injuries X is then

$$X = X_1 + X_2 + \ldots + X_N.$$

Note that

$$\mathbb{E}[X|N] = \mathbb{E}[X_1+\ldots+X_N|N] = N\cdot\mathbb{E}[X_i] = 2.5N.$$

Therefore

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|N]] = \mathbb{E}[2.5N] = 2.5\mathbb{E}[N] = 12.5.$$

Hence the expected number of injured workers in a week is 12.5.